基于 BLSTM 网络的医学时间短语识别 *

张顺利 [1]，王应军 [1]，姬东鸿 [2]

(1. 河南科技学院 信息工程学院，河南 新乡 453003; 2. 武汉大学 国家网络安全学院，武汉 430205)

摘 要：从医学文本中识别时间短语是临床医学自然语言处理的关键技术之一。传统基于规则和机器学习的方法，需要设计复杂规则和提取特征，而且大多数系统采用串行方法会导致错误的传播。提出基于双向长短时记忆网络（BLSTM）的神经网络架构，在识别时间表示式的同时判别它们的类型：首先使用卷积神经网络（CNN）学习得到单词的字符级别向量和大规模生物医学背景语料上训练得到的词向量进行组合作为 BLSTM 的输入，然后使用 BLSTM 网络学习单词的上下文语义表示，最后使用条件随机场（CRF）对 BLSTM 输出的序列进行标签优化。实验基于 SemEval-2016 任务 12，结果表明没有添加任何特征的神经网络学习方法比该任务中官方提供的最高分的 F1 值提高了 3%。

关键词：时间短语；病历文本；LSTM

中图分类号：TP391 doi: 10.19734/j.issn.1001-3695.2018.09.0742

Temporal phrases extraction in clinical text based on bidirectional long-short term memory model

Zhang Shunli[1], Wang Yingjun[1], Ji Donghong[2]

(1. *School of Information Engineering*, *Henan Institute of Science & Technology*, *Xinxiang Henan* 453003, *China*; 2. *National Network Security College*, *Wuhan University*, *Wuhan* 430205, *China*)

Abstract: Recognizing time phrases from clinical text is a fundamental task for many applications in clinical NLP. Previous work mainly adopted rule-based methods or machine learning approaches, which greatly depend on pattern designing or feature engineering, and the pipeline method used by most systems may lead to error propagation. This paper proposed a novel neural network based on bidirectional long-short term memory (BLSTM) to identifying clinical time expressions and the type of them simultaneously. First, it combined character-level word embedding trained by convolutional neural network (CNN) with word embedding trained from large-scale biomedical corpus together as input to BLSTM, then utilized BLSTM to model context information of each word, finally employed conditional random field (CRF) to optimize the output of BLSTM. This paper evaluated the model on the Semeval-2016 Task 12; it receives the best F1 value without requiring any handcrafted features or rules. Compared with the state-of-the-art systems in this task, the proposed model improves the F1 scores by 3%.

Key words: time expressions; clinical text; LSTM

病历文本中时间信息的抽取是近期生物医学研究的热点问题，这些信息可以用来帮助医务工作人员及科研工作者更好地认识疾病的发展模式，理解和认识动态的医学现象，是进行临床路径研究和智能决策支持系统开发的基础[1~3]。病历文本是医务人员对患者进行医疗活动过程的记录，由于医务人员工作的快节奏及专业性特点，病历文本中的时间信息存在着书写格式不统一、表述不规范以及和医学事件相互联系的特点，这些特点给时间短语的识别带来了很大的困难，阻碍了时间信息的后续利用，所以自动识别病历文本中的时间短语信息成为重要的研究课题，得到了越来越多的关注。

在英文病历文本中抽取时间信息，目前影响比较大的是系列 Clinical TempEval 共享任务，该任务已经在 Semeval 会议上举行了三次测评任务，分别为 Semeval2015 Task6 [4]、Semeval2016 Task12 [5]和 Semeval2017 Task12 [6]。该任务主要是从医院病历相关文本中抽取时间相关的信息，可以分为医学时间短语抽取，医学事件抽取、医学时间和事件关系抽取三个任务。医学时间短语识别是医学时间信息抽取任务中的第一步，也是最为关键的一步，它所识别的时间短语是医学时间和事件关系识别的基础，是整个任务的核心。

0 相关工作

从相关文献来看，目前时间短语的识别方法主要采用基于规则的方法和机器学习的方法。基于规则的识别方法认为，自然语言中基本的时间短语都有着清晰的结构和明显的特征，通过设计完备的规则可以覆盖大部分的时间信息。在通用领域，Strötgen 等人[7]采用基于正则表达式的模式设计了 HeidTime 系统来抽取时间表达式，然后再使用语言学的规则对表示式进行标准化转换，在 TempEval-2 的测评上取得了最高的 F1 值。Chang 等人[8]设计的 SUTIME 系统支持英文文本中四种时间类型的识别：时间（time）、持续的时间段（duration）、时间间隔（interval）和时间集合（set），现已经集成到斯坦福自然语言处理包中。Zhong 等人[9]提出了一个叫做 SynTime 的方法来抽取时间表达式，在这个方法在推特数据集上的识别效果超过了现有的最好方法。由于病历文本中医学时间短语的多样性和不确定性，采用通用领域的基于规则的方法需要花费巨大的工作量，例如 MayoTime[10]在

收稿日期：2018-09-11；**修回日期：**2018-11-16 **基金项目：**国家自然科学基金资助项目（61373108）；河南省重点科研研究项目（15A520069）

作者简介：张顺利（1981-），女，河南泌阳人，讲师，硕士，主要研究方向为自然语言处理、数据挖掘（zhangshunli@hist.edu.cn）；王应军（1980-），男，讲师，硕士，主要研究方向为自然语言处理、单片机应用；姬东鸿（1967-），男，教授，博士，主要研究方向为自然语言处理、信息检索．

2012 i2b2 的医学信息识别任务中设计了 300 多个规则完成了医学时间短语的抽取工作。

目前多数研究者采用机器学习中的线性统计模型来完成医学时间短语的抽取任务。在 Semeval-2016 的任务 12 中，大部分小组采用此类方法完成了时间短语的抽取任务。UTHealth 组[11]采用隐性马尔可夫（HMM）和支持向量机（SVM）相结合的序列标注器进行时间短语的识别，取得了最高的 F1 值。LIMSI 组[12]和 GUIR 组[13]使用基于条件随机场（CRF）的序列标注算法抽取时间短语。ULISBOA 组[14]使用基于支持向量机的序列标注器。以上这些小组无一例外地都使用了大量的句法、词法和领域知识作为特征，其中 UTHealth 组还使用了基于规则的系统 SUTIME[8]的识别结果作为特征。特征工程的构建需要花费大量的时间和精力，而且，在构建特征的过程中难免引入的一些不准确的特征，也会影响最终的识别效果。

近年来，循环神经网络的变体——基于双向长短时记忆单元的神经网络（BLSTM）由于能更好地学习到自然语言中的上下文语义特征，得到了广泛地应用。Huang 等人[15]抽取单词的拼写特征作为 BLSTM 网络的输入，然后将 BLSTM 网络的输出序列使用 CRF 模型的进行了输出优化，在 CONLL2003 的组块识别任务上取得了 88.83%的 F1 值。实验表明 BLSTM 模型可以有效得建模句子的上下文语义表示，CRF 可以对输出的标签优化，获得一个全局最优的标记序列。Ma 等人[16]使用 CNN 模型学习单词的字符级别向量表示，并组合词向量一起作为 BLSTM-CRF 模型的输入，在未使用任何人工特征的情况下在词性标注和通用命名实体识别任务上取得了不错的成绩，实验不仅验证了 BLSTM-CRF 在序列标注问题上的有效性，还说明了 CNN 模型可以有效地学习到单词的拼写特征。李丽双等人[17]使用 CNN-BLSTM-CRF 模型在 BiocreativeⅡGM 和 JNLPBA 语料上进行生物医学命名实体识别，取得了目前最好的成绩。在使用深度神经网络进行医学时间短语识别的研究中，Fries[18]在 SemEval-2016 任务 12 的测评任务中使用了一个简单的双向循环神经网络(vanilla 版本)进行医学时间短语的抽取，Chikka[19]使用卷积神经网络架构进行时间短语的抽取，识别效果比基于规则的方法和统计模型识别方法差。

从医学时间短语抽取流程的整体框架来看，大多数系统都是采用串行方法，即先识别出时间短语，再识别其类别，这种方法在一定程度上会造成错误的传播。Lee 等人[11]提出在识别时间表达式的同时识别其类型的方法，该方法一定程度上可以避免错误的传播。针对病历文本中时间表达的多样性、关联性和不确定性，作者提出了一个三层的基于双向长短时记忆单元的神经网络系统进行时间短语识别：a)使用 CNN 训练得到的字符级别向量和大规模生物医学背景语料上训练得到的词向量进行组合作为网络的第一层词向量输入层；b）第二层使用 BLSTM 网络学习单词的上下文语义表示；c）第三层使用 CRF 对 BLSTM 网络输出的序列进行标签输出优化。在 Semeval-2016 的任务 12 中语料上的实验结果表明，在未使用任何人工特征的情况下，该模型达到了目前的最好效果。

1　模型及算法描述

病历文本中的时间短语抽取任务可以看成序列标注任务，参考先前的成功案例，在识别时间表达式的同时判别出时间的类型。采用 BIO 方式进行结果标注。例如用“B-Timex-Date”来表示一个单词是 Date 类型的时间表达式的开头，用“I-Timex-Time”方式来表示一个单词是 Time 类型的时间表达式的非开头部分。

图 1 为本文提出的基于长短时记忆单元的神经网络架构。其共分为三层来实现：第一层为词向量表示层，第二层为双向长短期记忆层（BLSTM），第三层为标签输出层。首先将句子分离成一个单词序列，通过查询词向量表将输入的单词转换为相应的词向量序列，然后和使用卷积神经网络（CNN）训练出的单词的字符级别表示向量相拼接，拼接后的词向量用作双向长短时记忆层的输入，经过该层计算后输出带有单词上下文语义的向量表示，最后在标签输入层使用最大条件随机场（CRF）对输出的标签进行整体优化。

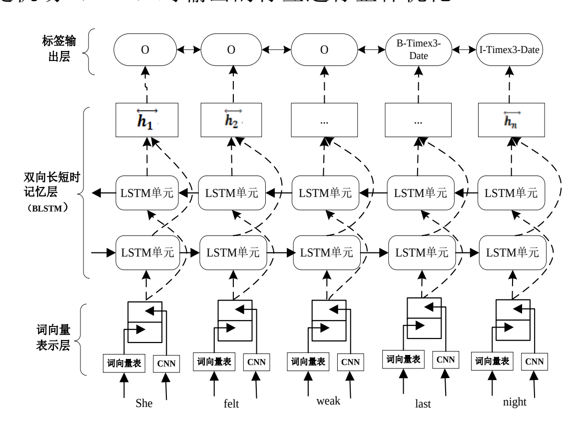


图 1　基于长短时记忆单元的神经网络架构

Fig. 1　Neural network architecture based on BLSTM

1.1　词向量表示层

词向量表示层的输入为单词，输出为该单词的向量表示。本研究中所使用的单词的向量表示由两部分向量拼接而成，一部分是由单词查询一个词向量表得到，另一部分是使用 CNN 网络计算的单词的字符级别表示向量。词向量表可以使用 word2vector [20]在大规模的非标注数据集上自己训练，也可以使用领域内提供的公开向量集，这种公共数据集在大规模的语料上通过非监督学习而得到，具有很好的泛化效果。

Ma 等人 [16]的研究表明，卷积神经网络可以通过卷积层能够很好地描述单词的局部特征，池化层可以进一步提取出单词局部特征中最具有代表性的部分。因此，本文提出使用 CNN 抽取病历文本中单词的字符级别特征（前缀、后缀以及拼写），再通过字符级别特征与词向量相结合的方法来提高模型的性能。本文使用的 CNN 模块与 Ma 和 Hovy 不同之处在于：本文在初始字符向量时，对字符的大小写、标点符号对应了不同的字符向量。

本文使用的 CNN 的结构如图 2 所示。对于一个长度为 n 的单词 $w=\{a_1,a_2,\cdots a_n\}$，a_i 为单词中的第 i 个字符，定义 $emb(a_i)$ 为该字符的字符向量，假设卷积神经网络的窗口大小 C 为 1，则字符 a_i 最终的向量表示为前后各一个字符和自身字符的向量的连接，记为 $r_{a_i}=[emb(a_{i-1}),emb(a_i),emb(a_{i+1})]$ 卷积操作使用一个固定大小的卷积核在单词的字符向量矩阵上进行卷积来提取出局部特征，最后使用最大池化操作获取整个单词的字符级别的特征向量表示。

1.2　双向长短期记忆层

双向长短时记忆（BLSTM）网络是一种特殊的循环神经网络（RNN）模型，它克服了传统 RNN 模型由于序列过长而产生的梯度消失问题[21,22]，已经被成功地运用到应用于实体识别、文本分类和机器翻译多个自然语言处理任务中。双向长短时记忆网络在结构上和双向 RNN 相似，唯一不同之处是它使用 LSTM 单元[23]替代了双向 RNN 网络上的隐藏单

元。LSTM 单元使用输入门、记忆单元、遗忘门、输出门来控制上下文信息被记忆或者被遗忘。其结构可以形式化地表示为

$$
\begin{aligned}
f_t &= \sigma(w_f \cdot [h_{t-1} \oplus x_t] + b_f) \\
i_t &= \sigma(w_i \cdot [h_{t-1} \oplus x_t] + b_i) \\
\tilde{c}_t &= \tanh(w_c \cdot [h_{t-1} \oplus x_t] + b_c) \\
c_t &= f_t \times c_{t-1} + i_t \times \tilde{c}_t \\
o_t &= \sigma(w_o \cdot [h_{t-1} \oplus x_t] + b_o) \\
h_t &= o_t \times \tanh(c_t)
\end{aligned} \tag{1}
$$

其中：σ 是激活函数 sigmod；i_t, f_t, o_t, c_t 分别表示在 t 时刻的输入门、忘记门、输出门和记忆门；w_i, w_f, w_c 代表权重矩阵；b_i, b_f, b_c 为偏置向量。对于一个输入的向量序列 $\{x_1, x_2, \cdots x_n\}, t \in 1 \cdots n$，每一个 LSTM 单元的输入为 $\{x_t, h_{t-1}, c_{t-1}\}$，输出为 $\{h_t, c_t\}$。i_t 决定哪些新的信息被存放在当前的记忆单元 c_t 中，f_t 决定哪些信息在当前的单元中将被丢弃掉，o_t 决定哪些信息将被输出到当前的 h_t 中。

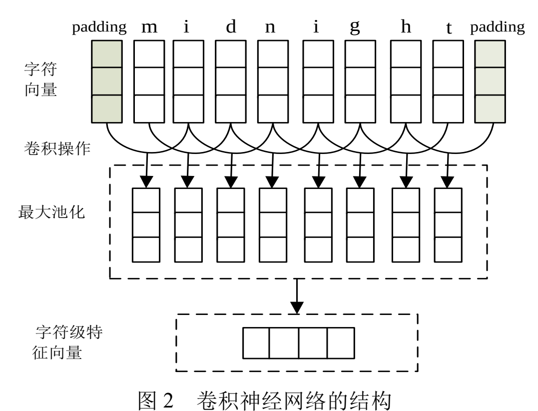


图 2 卷积神经网络的结构

Fig. 2 Structure of convolutional neural network

为了能够有效利用单词的上下文信息，本文采用如图 1 所示的 BLSTM 结构。BLSTM 网络对每个输入句子分别采用从左到右地顺序（forward）和从右到左地顺序（backward）计算，每一个句子中的第 t 个单词经过计算后得到两种不同的隐藏层向量表示：$h^{(l)}_t$ 和 $h^{(r)}_t$，双向长短时记忆层的输出由这两个向量的拼接计算得到，其公式如式（2）所示。

$$\vec{h}_t = \tanh(W_h \cdot (h^{(l)}_t \oplus h^{(r)}_t) + b_h) \tag{2}$$

1.3 标签输出层

条件随机场(CRF)，是由 Lafferty 等人[24]于 2001 年提出的无向概率图模型。CRF 通过考虑相邻标签的关系获得一个全局最优的标记序列，近几年在许多序列标注任务中都有很好的表现。本文使用 CRF 算法对 BLSTM 层的输出结果进行优化，获得全局最优的标签输出。

对于一个句子：

$$x = \{x_1, x_2, \cdots x_n\}$$

定义 P 为其在 BLSTM 层计算后输出的评分结果。P 为 n × k 的矩阵, k 为输出的标签种类个数，定义 $P_{i,j}$ 为句子中第 i 个单词输出第 j 个标签的概率。对于一个预测序列 $y = \{y_1, y_2, \cdots y_n\}$，它的评分可以定义为

$$S(x, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \tag{3}$$

其中：A 是转移矩阵；$A_{i,j}$ 代表由标签 i 转移到标签 j 这里使用 softmax 函数定义产生序列 y 的概率。

$$p(y|X) = \frac{e^{S(x,y)}}{\sum_{\tilde{y} \in Y_X} e^{S(x,\tilde{y})}} \tag{4}$$

训练数据的似然函数为

$$\log(p(y|X)) = S(x, y) - \log\left(\sum_{\tilde{y} \in Y_X} e^{S(x,\tilde{y})}\right) \tag{5}$$

Y_X 表示对应一个句子 X 所有可能的标记序列。通过上面的公式得到一个有效合理的输出序列。预测时，由公式输出整体概率最大的一组序列：

$$y^* = \mathrm{argmax}(S(x, \tilde{y})) \tag{6}$$

1.4 训练参数

训练过程中，采用随机梯度下降模型 AdaGrad [25],学习率选取为 0.03，正则化参数为 10^{-8}。为了减轻模型的过度拟合，在 BLSTM 层的输入/输出部分增加了 Dropout，其值定为 0.5。参数根据开发集的结果上进行了调整。本文选取可公共使用的 200 维的 Pubmed[26]向量集作为单词的初始向量查询表，此向量集从大量的生物医学文献和摘要中训练而成。向量集选取的实验依据将在 3.2.1 节中给出。其他向量定义为 200 维。CNN 的窗口定为 1，输出向量长度定为 20。

2 实验

2.1 数据集与评测方法

文中采用 Semeval-2016 任务 12 提供的梅奥医学中心标注的癌症病人的病理报告和临床记录语料(THYME 语料)[27]进行了各项实验。数据集分为训练集、开发集和测试集三部分。其中训练集中含有 293 篇临床记录文档，开发集和测试数据集分别含有 147 篇和 152 篇文档，它们分别包含 3 833、3 078、1 952 个时间表达式。

在实验中使用和该任务一致的评测方法，在下面的表中 P 代表准确率(precession),R 代表召回率（recall），F1 代表 F1 值（F1-measure）。

2.2 实验结果

2.2.1 词向量的表示

词向量的表示对序列标注任务的结果有着很大的影响，文中选取以下三类词向量进行了对比实验：a）随机初始化长度为 100 维、200 维、300 维的词向量进行实验；b）采用 word2vec 工具，对任务提供的 THYME 语料进行训练得到的 100 维、200 维、300 维的词向量；c）由 Pyysalo[26]从大规模生物医学语料中训练出来的 200 维的 Pubmed 词向量集。实验结果如表 1 所示，第三类向量取得了最好的泛化结果。

表 1 词向量的选取对识别性能的影响

Table 1 Results with different choices of word embedding

词向量表示	P	R	F1
随机初始 100 维	81.75	78.31	79.99
随机初始 200 维	82.35	78.68	80.47
随机初始 300 维	82.36	78.43	80.35
THYME 语料训练 100 维	81.93	78.84	80.36
THYME 语料训练 200 维	82.22	79.84	81.01
THYME 语料训练 300 维	82.41	79.75	81.06
Pubmed200 维向量	83.72	80.03	81.83

2.2.2 神经网络架构分层实验

对神经网络结构进行分层测试，对比实验的结果来分析各个模块在模型中起到的作用。实验结果如表 2 所示。“-CRF”表示模型去掉 CRF 层，标签结果采用 softmax 进行输出。“-pretrain”表示模型不使用训练好的 Pubmed 向量集作为初始向量，而是采用随机方法初始词向量；“-CNN”表示模型去掉 CNN 层，从 Pubmed 向量集中读取的向量直接作为词向量的表示；“CNN-BLSTM-CRF”为本文实验最终所选用的神经网络模型。

表 2 模型的分层测试

Table 2 Ablation test performance evaluation

模型	P	R	F1
-CNN	80.54	79.82	80.18
-pretrain	82.35	78.68	80.47
-CRF	79.86	78.34	79.09
CNN-BLSTM-CRF	83.72	80.03	81.83

2.2.3 Dropout 设置

Dropout 可以防止神经网络模型的过度拟合。文中采用不同的 Dropout 数值，在数据集上进行了测试，结果如表 3 所示。

表 3 不同 Dropout 对模型性能的影响

Table 3 Results with different dropout values

Dropout	P	R	F1
-	82.51	79.38	80.91
-0.1	82.82	79.63	81.19
-0.3	83.29	79.92	81.57
-0.5	83.72	80.03	81.83
-0.8	83.16	79.92	81.51

2.3 与现有其他工作的对比

经过上述三组实验，文中选取维度为 200 的 Pubmed 向量，Dropout 为 0.5 进行实验，并将实验结果与其他优秀学者的实验结果相比较，比较结果如表 4 所示。

表 4 和其他学者的工作比较

Table 4 Performance comparison with previous research

团队	P	R	F1
CDE-IIITH (Chikka, 2016)*	0.614	0.560	0.586
Brundlefly (Fries, 2016)*	68.60	41.50	51.70
UFPRSheffield(Tissot et al., 2015)	0.311	79.50	44.70
UTHealth (Lee et al., 2016)	83.60	75.70	79.50
LIMSI-1 (Grouin and Moriceau, 2016)	84.00	51.00	63.50
CNN-BLSTM-CRF(文中模型)	83.72	80.03	81.83

Lee 等人[11]使用隐性马尔可夫（HMM）与支持向量机（SVM）相结合的序列标注器，结合大量的特征工程（词形、词性、词干和相关字典等）进行时间短语的识别，在 Semeval-2016 的评比中取得了最好的 F1（79.5）值。Grouin 等人[12]使用条件随机场（CRF）和基于规则的系统（HeidelTime）相结合的方法取得了 Semeval-2016 任务评比中准确率第一名，但是 F1 值却非常不理想。Tissot 等人[28]设计一个基于规则的医学时间识别系统，在 Semeval-2015 的测评中取得了最好的成绩。在使用神经网络的方法中，Fries[18]使用在两个医学语料上训练得到的词向量输入双向循环神经网络(vanilla 版本)进行医学时间短语的抽取，准确率和 F1 值均很低。Chikka[19]使用卷积神经网络架构进行了时间短语的抽取实验，结果低于其用 SVM 获得的识别效果。

通过以上对比分析可以看出，本文的神经网络模型在未使用任何人工特征的情况下，取得了目前最好的成绩。

2.4 错误分析

对实验的结果进行分析，主要发现有如下两类错误：

a）时间短语中的介词经常没有被识别出来，例如标准答案中的时间短语“for the past twenty years”，本文识别的短语为“the past twenty years”，漏掉了介词“for”。在训练语料中，一些介词时间短语的标注标准中有的包含介词，有的不包含介词，这会影响最终的识别效果。

b）样本的数量会影响神经网络系统的识别效果，例如 Time 类型的短语在训练数据中所占的比例很少，识别的正确率也相应地非常低。在神经网络的训练过程中，如何解决由于样本的不平衡而引起的结果偏差，也是本文要解决的一个问题。

3 结束语

本文提出了一个基于深度神经网络架构的病历时间短语抽取模型，模型使用卷积神经网络有效地表示了单词的词形特征，通过 BLSTM 网络序列上下文语义信息，并使用 CRF 算法对标签输出结果进行了优化，在没有使用任何人工特征和医学领域背景知识的情况下性能优于目前最好的系统。模型还可以用来解决一些类似的序列标注问题，如进行医学事件的抽取。将来，可以通过多任务的联合学习，使系统获得更好地泛化能力。譬如，本文可以训练一个同时识别时间和事件信息的联合神经网络模型来获得更好地效果。

参考文献：

[1] Hao Tianyong, Pan Xiaoyi, Gu Zhiying, *et al.* A pattern learning-based method for temporal expression extraction and normalization from multi-lingual heterogeneous clinical texts [J]. Bmc Medical Informatics & Decision Making, 2018, 18 (Suppl 1): 22.

[2] Moharasar G, Tu B H. A semi-supervised approach for temporal information extraction from clinical text [C]// Proc of IEEE Rivf International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future. 2016: 7-12.

[3] Liu Zengjian, Tang Buzhou, Wang Xiaolong, *et al.* CMedTEX: a rule-based temporal expression extraction and normalization system for chinese clinical notes [C].//Proc of AMIA Annual Symposium Proceedings. 2017: 818.

[4] Bethard S, Derczynski L, Savova G, *et al.* SemEval-2015 task 6: clinical tempeval [C]// Proc of the 9th International Workshop on Semantic Evaluation. 2015: 806-814.

[5] Bethard S, Savova G, Chen Weite, *et al.* Semeval-2016 task 12: clinical tempeval [C]// Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1052-1062.

[6] Bethard S, Savova G, Palmer M, *et al.* SemEval-2017 task 12: clinical tempeval [C]// Proc of the 11th International Workshop on Semantic Evaluation. 2017: 565-572.

[7] Strötgen J, Gertz M. HeidelTime: high quality rule-based extraction and normalization of temporal expressions [C]// Proc of International Workshop on Semantic Evaluation. 2010: 321-324.

[8] Chang A X, Manning C D. SUTIME: a library for recognizing and normalizing time expressions [J]. Lrec, 2012, 9 (1): 3735-3740.

[9] Zhong Xiaoshi, Sun Aixin, Cambria E. Time expressions analysis and recognition using syntactic token types and general heuristic rules [C]// Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver Canada: ACL, 2017: 420-429.

[10] Sohn S, Wagholikar K B, Li Dingcheng, *et al.* Comprehensive temporal information detection from clinical text: medical events, time, and tlink identification [J]. Journal of the American Medical Informatics Association Jamia, 2013, 20 (5): 836-842.

[11] Lee H J, Xu Hua, Wang Jingqi, *et al.* UTHealth at SemEval-2016 task 12: an end-to-end system for temporal information extraction from clinical notes [C]// Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1292-1297.

[12] Grouin C, Moriceau V. LIMSI at SemEval-2016 task 12: machine learning and temporal information to identify clinical events and time expressions [C]// Proc of the 10th International Workshop on Semantic

Evaluation. 2016: 1225-1230.

[13] Cohan A, Meurer K, Goharian N. GUIR at SemEval-2016 task 12: temporal information processing for clinical narratives [C]// Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1248-1255.

[14] Barros M, Lamurias A, Figueiro G, *et al.* ULISBOA at SemEval-2016 task 12: extraction of temporal expressions, clinical events and relations using ibent [C]// Proc of the 10th International Workshop on Semantic Evaluation. 2016.

[15] Huang Zhiheng, Xu Wei, Yu Kai. Bidirectional lstm-crf models for sequence tagging [J]. Computer Science, 2015.

[16] Ma Xuezhe, Hovy E. End-to-end sequence labeling via bidirectional lstm-cnns-crf [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: Association for Computational Linguistics. 2016: 1064-1067.

[17] 李丽双，郭元凯. 基于 CNN-BLSTM-CRF 模型的生物医学命名实体识别 [J]. 中文信息学报，2018, 32 (1): 117-122. (Li Lishuang, Guo Yuankai. Biomedical name entity recognition with cnn-blstm-crf [J]. Journal of Chinese Information Processing, 2018, 32 (1): 117-112.)

[18] Fries J. Brundlefly at SemEval-2016 task 12: Recurrent neural networks vs. joint inference for clinical temporal information extraction [C]// Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1274-1279.

[19] Chikka V R. CDE-IIITH at SemEval-2016 task 12: extraction of temporal information from clinical documents using machine learning techniques [C]// Proc of the 10th International Workshop on Semantic Evaluation. 2016: 1237-1240.

[20] Mikolov T, Sutskever I, Chen Kai, *et al.* Distributed representations of words and phrases and their compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.

[21] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Trans on Neural Networks, 1994, 5 (2): 157-166.

[22] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6 (2): 107-116.

[23] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional lstm and other neural network architectures [J]. Neural Networks, 2005, 18 (5): 602-610.

[24] Lafferty, John D, McCallum, *et al.* Conditional random fields: probabilistic models for segmenting and labeling sequence data [M]// Departmental Papers (CIS) . 2001: 282-289.

[25] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization [J]. Journal of Machine Learning Research, 2011, 12 (7): 257-269.

[26] Pyysalo S, Ginter F, Moen H, *et al.* Distributional semantics resources for biomedical text processing [C]// Proc of LBM. 2013: 39-44.

[27] 4Th S W, Bethard S, Finan S, *et al.* Temporal annotation in the clinical domain [J]. Trans of the Association for Computational Linguistics, 2014, 2 (1): 143-154.

[28] Tissot H, Gorrell G, Roberts A, *et al.* UFPRSheffield: contrasting rule-based and support vector machine approaches to time expression identification in clinical tempeval [C]// Proc of the 9th International Workshop on Semantic Evaluation. 2015: 835-839.